

5 CLAIMS:

What is claimed is:

1. A method for determining the similarity of data records in first and second data sets, the data records having an informational content, the method comprising:
identifying a first data record in the first data set that is potentially identical to a
10 second data record in the second data set, the identified first and second data records having an informational content that is non-identical but similar;
determining whether the first and second data records identified as potentially identical are truly identical based upon a predetermined criteria.
- 15 2. The method of claim 1 wherein identifying a first and second data records identifies telecommunication call detail records (CDRs).
3. The method of claim 1 wherein identifying a first data record and a second data record includes grouping the records in the first and second data sets into groups
20 based upon a predetermined criteria.
4. The method of claim 1 wherein identifying includes comparing the informational content of first data record to the informational content of the second data record.
25
5. A method for determining different data records in a telecommunications system from records in first and second data sets in a comprising:
identifying potentially different data records in the first data set at least in part by comparing records in the first data set to records in the second data set; and
30 verifying that the potentially different records identified as potentially different are truly different using at least one predetermined criteria.
6. The method of claim 5 wherein the different data records can be faulty data records or mismatched data records.
35

5 7. The method of claim 5 wherein determining potentially different data records includes defining a set of similarity characteristics, grouping the data records in each of the first and second sets according to the similarity characteristics into similarity groups, and comparing the similarity groups in the first data set to the similarity groups in the second data set.

10

 8. The method of claim 5 wherein determining potentially different data records includes determining whether each record in the first data set completely matches with a data record in the second data set.

15

 9. The method of claim 5 wherein verifying includes determining from the second data set a set of data records that are similar to the potentially different record identified in the first data set.

20

 10. The method of claim 5 wherein verifying includes scoring elements of each of the plurality of data records to form a plurality of scores, multiplying the plurality of scores to form a test score, comparing the test score to a predetermined minimum score, and determining a different record if the comparison determines the test score is unacceptable.

25

 11. The method of claim 5 further comprising taking an action relating to the different data records.

30

 12. A device for determining faulty data records in a telecommunications system from records in first and second data sets in a comprising:

 a data store containing first and second data sets; and

 a processor coupled to the data store and having an output,

 such that the processor identifies potentially different data records in the first data set at least in part by comparing records in the first data set to records in the second data set and verifies that the potentially different records identified as potentially different are

5 different using at least one predetermined criteria and identifies the different records on the output.

13. The device of claim 12 wherein the processor includes means for defining a set of similarity characteristics, means for grouping the data records in each of the first
10 and second sets according to the similarity characteristics into similarity groups, and means for comparing the similarity groups in the first data set to the similarity groups in the second data set.

14. The device of claim 12 wherein the processor includes means for
15 determining whether each record in the first data set completely matches with a data record in the second data set.

15. The device of claim 12 wherein the processor includes means for
20 determining from the second data set a set of data records that are similar to the potentially faulty record identified in the first data set.

16. The device of claim 12 wherein the processor includes means for scoring elements of each of the plurality of data records to form a plurality of scores, means for multiplying the plurality of scores to form a test score, means for comparing the test score
25 to a predetermined minimum score, and means for determining a different record if the comparison determines the test score is unacceptable.

17. The device of claim 12 wherein the different data records can be faulty data records or mismatched data records.